PATENT
29250-000494/US

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

APPLICANTS:	Peng J. ZHANG	CONF. NO.:	3637
SERIAL NO.:	09/615,945	GROUP:	2626
FILED:	July 13, 2000	EXAMINER:	Michael N. Opsasnick
FOR:	METHOD AND APPARATUS FOR DISCRIMINATING SPEECH FROM VOICE-BAND DATA IN A COMMUNICATION NETWORK		

APPELLANT'S BRIEF ON APPEAL UNDER 37 C.F.R. § 41.37

Customer Service Window
Randolph Building
401 Dulany Street
Alexandria, VA 22314
Mail Stop Appeal Briefs - Patents

February 16, 2007

Sir:

In accordance with the provisions of 37 C.F.R. §41.37, Appellant submits the following:

TABLE OF CONTENTS

I. REAL PARTY IN INTEREST:

The real party in interest is Alcatel-Lucent.

II. RELATED APPEALS AND INTERFERENCES

No related appeals or interferences are known.

III. EVIDENCE SUBMITTED UNDER 37 C.F.R. 1.130, 1.131 OR 1.132:

None.

IV. DECISIONS RENDERED BY A COURT OR THE BOARD IN RELATED APPEALS AND INTERFERENCES SECTION:

None.

V. STATUS OF CLAIMS:

Claims 1-21 are pending; with claims 1 and 11 being written in independent form.

Claims 1-21 stand finally rejected under 35 U.S.C. § 103(a) as allegedly being unpatentable over Sewall (U.S. Patent 6,708,146) in view of Huang (U.S. Patent 6,018,706)

Claims 1-21 are being appealed.

VI. STATUS OF AMENDMENTS:

An Amendment After Final was filed and entered on October 23, 2006.

VII. SUMMARY OF CLAIMED SUBJECT MATTER:

A. Concise explanation of the subject matter set forth in each of independent claims 1 and 11, and each dependent claim (e.g., dependent claims 5-9) argued separately.

i. The following explains the subject matter set forth in each independent claim and each dependent claim argued separately referring to the specification by page and line number, and to the drawings, if any, by reference characters in accordance with 37 C.F.R. § 41.37(c)(1)(v).

Claim 1 is directed to a method of discriminating speech from voice-band data (VBD) in a communication network (*see, e.g.*, FIG. 1 and FIG. 2; "Speech" and "VBD"). The method of claim 1 comprises calculating a self similarity ratio value representing a periodicity characteristic (*e.g.*, measure similarity between sequential signal segments of an input signal, *see, e.g.*, element 150 of FIG. 2; element 208 of FIG. 3A; element 238 of FIG. 3B; element 268 of FIG. 3C; p. 6, ll. 20-25 and p. 7, ll. 1-8, equations 3, 4, and 5). The method of claim 1 further comprises calculating an autocorrelation coefficient value representing a spectral characteristic (*e.g.*, analyze the spectral characteristics of a frame of interest, *see, e.g.*, element 150 of FIG. 2; element 208 of FIG. 3A; element 238 of FIG. 3B; element 268 of FIG. 3C; p. 7, ll. 16-26 and p. 8, ll. 1-14; "R2d", "R3d", "R4d"). As set forth in the method according to claim 1, calculating the self similarity ratio value includes calculating a plurality of different self similarity ratio values (*see, e.g., p. 7, "COL (n, j)" of equation 4*) and selecting the highest one of the plurality of different self similarity ratio values as the calculated self similarity ratio value (*see, e.g.*, p. 6, l. 20 – p. 7, l. 8; "SSR1" and "SSR2" of equations 3 and 5). The method of claim 1 further comprises determining whether said input signal segment is speech or voice-band data (*see, e.g.*, FIG. 2, "SPEECH" and "VBD") based on said at least one of said self similarity value and

said autocorrelation coefficient value (*see, e.g.*, elements 152, 158, 162, and 166 of FIG. 2; element 210 of FIG. 3A; element 240 of FIG. 3B; element 270 of FIG. 3C).

Claim 11 is directed to apparatus (*see, e.g.*, unit 130 of FIG. 1) for discriminating speech from voice-band data in a communication network (*see, e.g.*, FIG. 1 and FIG. 2; "Speech" and "VBD"). The apparatus of claim 11 comprises calculating means for calculating a self similarity ratio value, representing a periodicity characteristic (*e.g.*, measure similarity between sequential signal segments of an input signal, *see, e.g.*, element 150 of FIG. 2; element 208 of FIG. 3A; element 238 of FIG. 3B; element 268 of FIG. 3C; p. 6, ll. 20-25 and p. 7, ll. 1-8, equations 3, 4, and 5), and an autocorrelation coefficient value, representing a spectral characteristic (*e.g.*, analyze the spectral characteristics of a frame of interest of an input signal, *see, e.g.*, element 150 of FIG. 2; element 208 of FIG. 3A; element 238 of FIG. 3B; element 268 of FIG. 3C; p. 7, ll. 16-26 and p. 8, ll. 1-14; "R2d", "R3d", "R4d"). As set forth in claim 11, calculating the self similarity ratio value includes calculating a plurality of different self similarity ratio values (*see, e.g.*, p. 7, "COL (n, j)" of equation 4) and selecting the highest one of the plurality of different self similarity ratio values as the calculated self similarity ratio value(*see, e.g.*, p. 6, l. 20 – p. 7, l. 8; "SSR1" and "SSR2" of equations 3 and 5). The apparatus of claim 11 further comprises determining means for determining whether said input signal segment is speech or voice-band data (*see, e.g.*, FIG. 2, "SPEECH" and "VBD") based on said at least one of said self similarity value and said autocorrelation coefficient value (*see, e.g.*, elements 152, 158, 162, and 166 of FIG. 2; element 210 of FIG. 3A; element 240 of FIG. 3B; element 270 of FIG. 3C).

In the method of claim 5, said calculating step calculates a first autocorrelation coefficient as a first spectral characteristic value (*e.g.*,

analyze the spectral characteristics of a frame of interest of an input signal, *see, e.g.*, element 150 of FIG. 2). The method of claim 5 further sets forth that said determining step determines that said input signal segment is voice-band data if said first autocorrelation coefficient is less than a first autocorrelation threshold (*see, e.g.*, FIG. 2 element 152, "VBD"), and that said input signal segment is speech if said first autocorrelation coefficient is greater than a second autocorrelation threshold (*see, e.g.*, FIG. 2 element 152, "SPEECH"), said second autocorrelation threshold being greater than said first autocorrelation threshold (*see, e.g.*, FIG. 2, "TR2H" and "TR2L"; *see, e.g.*, p. 9, ll. 19-27).

In the method of claim 6, said calculating step calculates second and third autocorrelation coefficients as second and third spectral characteristic values respectively (*e.g.*, analyze the spectral characteristics of a frame of interest of an input signal, *see, e.g.*, element 150 of FIG. 2; and "R3D", "R4D" of FIG. 2), and said determining step determines that said input signal segment is voice-band data if said second autocorrelation coefficient is less than a third autocorrelation threshold or said third autocorrelation coefficient is less than a fourth autocorrelation threshold (*see, e.g.*, element 162 of FIG. 2 and "TR3", "TR4" of FIG. 2; *see, e.g.*, p. 9, l28 – p. 10, l. 9).

In the method of claim 7, said determining step determines that said input signal segment is voice-band data if a sum of said second autocorrelation coefficient and said third autocorrelation coefficient is less than a fifth autocorrelation threshold (*see, e.g.*, element 162 of FIG. 2 and "R3d + R4d ≤ TR34" of FIG. 2).

In the method of claim 8, said calculating and determining steps are performed for a plurality of input signal segments (*e.g.*, frames or segments of an input signal; *see, e.g.*, FIG. 3A, "NEXT FRAME") in accordance with a sequential decision logic sequence (*see, e.g.*, FIG. 3A,

3B, and 3B; p. 10, ll. 19-21) which designates input signal segments as speech during a speech state (*see, e.g.*, elements 230 and 240 of FIG. 3B) and designates input signal segments as voice-band data during a voice-band data state (*see, e.g.*, elements 260 and 270 of FIG. 3C).

In the method of claim 9, said sequential decision logic sequence switches from said speech state (*see, e.g.*, "B" and elements 230 and 248 of FIG. 3B) to said voice-band data state (*see, e.g.*, "C" and element 260 of FIG. 3C) when results of said determining step for a plurality of input signal segments indicate that said speech state is erroneous (*see, e.g.*, "VBD" and elements 240, 244, 246, and 248 of FIG. 3B). Furthermore, in the method of claim 9, said sequential decision logic sequence switches from said voice-band data state (*see, e.g.*, "C" and elements 260 and 278 of FIG. 3C) to said speech state (*see, e.g.*, "B" and element 230 of FIG. 3B) when results of said determining step for a plurality of input signal segments indicate that said voice-band data state is erroneous (*see, e.g.*, "SPEECH" and elements 270, 274, 276, and 278 of FIG. 3C).

ii. The following is a more general discussion of the subject matter of the application to assist the board of appeals in understanding example embodiments of the present invention.

Generally, independent claim 1 reads on methods for discriminating speech from voice-band data in a communication network as illustrated, for example, in FIG. 2. Independent claim 11 reads on apparatuses for discriminating speech from voice-band data in a communication network as illustrated, for example, in FIG. 1 and 2. Dependent claims 5-9 read on methods of discriminating speech from voice-band data in a communication network as illustrated, for example, in FIG. 3A, 3B, and 3C. Such is discussed in more detail below.

Typically, voiced speech is characterized by relatively high energy content and periodicity, i.e., "pitch", while unvoiced speech exhibits little

or no periodicity. Transition regions which occur between voiced and unvoiced speech regions often have characteristics of both voiced and unvoiced speech. During normal transmission, high-speed voice-band date (VBD) is scrambled, encoded, and modulated, thereby appearing as noise with no periodicity. Some low-speed VBD signals, such as control signals used during a start-up procedure, exhibit periodicity. The present invention discriminates between periodic speech and VBD signals by recognizing that periodic VBD signals will typically have a faster repetition rate than voiced speech, and calculating short-term delay and long-term delay self similarity ratio (SSR) values to indicate the repetition rate of an input signal frame.

The present invention also recognizes that analyzing the periodicity characteristics of an input frame may not ensure accurate speech/VBD discrimination, and that the certain spectral characteristics of an input frame may reveal whether the input frame is speech or VBD. For example, the carrier frequency used by a typical modem/fax is within a narrow range, whereas speech is a non-stationary random signal that typically exhibits large variations in its power spectrum. The present invention calculates short-term autocorrelation coefficients to determine the spectral envelope of an input frame to facilitate accurate speech/VBD discrimination.

Fig. 1 of the subject application is a general block diagram of an apparatus for discriminating speech from VBD signals in accordance with one embodiment of the present invention. As shown in Fig. 1, the speech/VBD discriminator 100 includes an input frame buffer 110, a high-pass filter 120, and a speech/VBD discriminating unit 130. The input frame buffer 110 receives an input signal (e.g., from a network line card which samples the signal from a conventional telephone network

channel at an 8 kHz clock rate) to buffer frames of *N* consecutive speech samples per frame. Nominally, the input signal received by the input frame buffer has been sampled at an 8 kHz clock rate, frame size is in the range of 10 milliseconds (i.e., $N = 80$ samples at a 8 kHz sampling rate) to 30 milliseconds (i.e., $N = 240$ samples at a 8 kHz sampling rate), and a 16-bit linear binary word represents the amplitude of an input sample (i.e., an input sample is no more than 2^{15}). The high-pass filter 120 filters each frame of *N* samples to remove DC components therefrom.

The speech/VBD discriminating unit 130 calculates short-time power, *Ps*, of an input frame using a window of *N* samples. The speech/VBD discriminating unit 130 also calculates SSR values to measure the similarity between sequential signal segments (SSR1 and SSR2). The speech/VBD discriminating unit 130 also calculates autocorrelation coefficients (R2d, R3d, and R4d), which represent certain spectral characteristics of the frame of interest.

Fig. 2 of the subject application illustrates a "raw decision" sequence for classifying a single input frame as being either speech or VBD. After calculating the *Ps*, *SSR1*, *SSR2*, *R2d*, *R3d*, and *R4d* values discussed above (step 150), the speech/VBD discriminating unit 130 initially attempts to classify the frame of interest as either speech or VBD based on *R2d* (step 152).

If *R2d* is between *TR2L* and *TR2H*, then the speech/VBD discriminating unit 130 next attempts to achieve a discrimination result based on *SSR1* (step 158). If *SSR1* is less than *TS1*, the speech/VBD discriminating unit 130 next attempts to discriminate based on *R3d* and *R4d* (step 162).

If none of these conditions are met, the speech/VBD discriminating unit 130 next attempts to discriminate based on *SSR2* (step 166).

The speech/VBD discrimination technique described above is implemented in a sequential decision logic algorithm in accordance with one embodiment of the present invention to improve decision reliability.

Figs. 3A-3C of the subject application are flowcharts which illustrate an exemplary sequential decision logic algorithm implemented by the speech/VBD discriminating unit 130 to discriminate speech and VBD. The sequential decision logic algorithm illustrated in Figs 3A-3C essentially has six states: (1) an initialization state; (2) a determination state in which individual input frames are classified as being either speech or VBD; (3) a speech state in which the classification result remains speech until subsequent classification results indicate that the speech state is erroneous; (4) a "was speech" state in which a period of low-power occurs after entering the speech state; (5) a VBD state in which the classification result remains VBD until subsequent classification results indicate the VBD state is erroneous; and (6) a "was VBD" state in which a period of low-power occurs after entering the VBD state. The significance of these classification states will become more apparent from the following description.

Referring to Fig. 3A of the subject application, during an initialization step, each counter used in the sequential decision algorithm is set to 0 (step 202). Next, the discriminating unit 130 calculates *Ps* for a frame of interest (step 204) and determines whether *Ps* is greater than or equal to an energy threshold *ETh1* (step 206). When *Ps* is less than *ETh1*, the discriminating unit does not attempt to determine whether the frame is speech or VBD, and instead returns to step 204 to calculate the *Ps* for the next frame. In other words, the discriminating unit 130 does

not initially attempt to classify input frames as speech or VBD until *Ps* reaches *ETh1*. The sequential decision logic algorithm remains in an initialization state until *Ps* reaches *ETh1*.

When the discriminating unit 130 determines that *Ps* is greater than or equal to *ETh1*, the sequential decision logic algorithm enters a determination state in which the speech/VBD discriminating unit 130 calculates discrimination feature values for the frame of interest (step 208) and decides whether these discrimination feature values indicate that the frame of interest is speech or VBD (step 210). In other words, the discriminating unit 130 executes the raw decision logic discussed above with reference to Fig. 2 to classify the frame of interest as speech or VBD. When the frame of interest is classified as speech, a speech counter *Spc* is incremented by 1 (step 212), and *Spc* is compared to a speech count threshold *Spy*, e.g., *Spy* = 1 (step 214). If *Spc* is less than *Spy*, the sequential decision logic remains in the determination state and the discriminating unit 130 computes the discrimination feature values for the next input frame (step 208). If *Spc* is at least equal to *Spy*, the sequential decision logic enters the speech state, which is described below with reference to Fig. 3B.

If, at step 210, the input frame is classified as VBD, a VBD counter *Mdc* is incremented by 1 (step 216), and *Mdc* is compared to a VBD count threshold *Mdy, e.g., Mdy* = 4. If *Mdc* is less than *Mdy*, the sequential decision logic remains in the determination state, and the discriminating unit 130 computes the discrimination feature values for the next frame (step 208). If *Mdc* is at least equal to *Mdy*, the sequential decision logic enters the VBD state, which is discussed in detail below with reference to Fig. 3C. In accordance with the sequential decision logic shown in Fig. 3B, after a predetermined number of frames have been classified as speech/VBD based on SSR and/or autocorrelation coefficient values so

that the sequential decision logic algorithm enters the speech/VBD state, speech/VBD discrimination output does not change unless a certain number of subsequent classification results indicate that the speech/VBD state is erroneous.

Referring to Fig. 3B of the subject application, when the sequential decision logic enters the speech state (step 230), *Ps* is calculated for the next frame (step 204) and compared with the energy threshold *ETh1* (step 234). If *Ps* is at least equal to *ETh1*, a silence counter *Sic* is set equal to 0 (step 236), and the speech/VBD discriminating unit 130 calculates discrimination feature values for the next frame (step 238) so that the input frame can be classified as speech or VBD (step 240), i.e., "raw decision" is performed. If the input frame is classified as speech at step 240, the VBD counter *Mdc* is divided by 2 (step 242), the sequential decision logic remains in the speech state, and the classification sequence returns to step 232 so that the discriminating unit 130 calculates *Ps* for the next frame. If the input frame is recognized as VBD at step 240, the VBD counter *Mdc* is incremented by a "power-compensated" increment *x* (described in detail below) (step 244), and *Mdc* is compared with the VBD state-change threshold *Mdx*, e.g., *Mdx* = 8 (step 246). If *Mdc* is not at least equal to *Mdx*, the sequential decision logic remains in the speech state, and the decision sequence returns to step 232 so that the speech/VBD discriminating unit 130 calculates *Ps* for the next frame. When, however, *Mdc* is at least equal to *Mdx*, the VBD counter *Mdc* is reset to 0 (step 248), and the sequential decision logic switches to the VBD state.

When the speech/VBD discriminating unit 130 determines at step 234 that *Ps* is less than *ETh1*, the silence counter *Sic* is incremented by 1 (step 250) and compared to a silence counter threshold *Siy*, e.g., *Siy* =8, (step 252). If *Sic* has not reached *Siy*, the sequential decision logic

remains in the speech state, and proceeds to step 238 so that the discriminating unit 130 computes discrimination values for the frame of interest. When *Sic* reaches *Siy*, however, the sequential decision logic enters a "was speech" state which will next be described with reference to flow diagram blocks 253-257. During the "was speech" state, the discriminating unit 130 initially calculates *Ps* for the next frame (step 253), and compares *Ps* with the energy threshold *ETh1* (step 254). If *Ps* is greater than or equal to *ETh1*, the silence counter *Sic* is reset to 0 (step 255) and the sequential decision logic returns to the speech state in step 238. When the discriminating unit 130 determines that *Ps* is less than *ETh1* at step 254, the silence counter *Sic* is incremented by 1 (step 256) and *Sic* is compared to a second silence counter threshold *Six* (step 257), e.g., *Six* = 200. If *Sic* has not reached *Six*, the sequential decision logic remains in the "was speech" state, and *Ps* is calculated for the next frame at step 253. When *Sic* reaches *Six*, however, the sequential decision logic returns to its initialization state at step 202, i.e., reset occurs.

Referring next to Fig. 3C of the subject application, it can be seen that the sequential decision logic operates during the VBD state in a similar manner to the speech state described above with regard to Fig. 3B. As such, additional discussion will be omitted herein for the sake of brevity. Associated discussion and a more detailed description of FIG. 3C may be found in the Specification, starting from page 13 continuing onto page 16.

VIII. GROUNDS OF REJECTION TO BE REVIEWED ON APPEAL:

Appellant seeks the Board's review of the rejection of claims 1-21 as being obvious under 35 U.S.C. §103(a) over Sewall (U.S. Patent 6,708,146) in view of Huang (U.S. Patent 6,018,706).

IX. ARGUMENTS:

A. Appellant traverses the rejection of claims 1-21 under 35 U.S.C. §103(a) as being upatentable over Sewall (U.S. Patent 6,708,146) in view of Huang (U.S. Patent 6,018,706).

The claims will be argued in 6 separate groups, to aid in the board's review of the associated arguments. Group 1 consists of claims 1, 2-4, 10, 11, and 12-21. Group 2 consists of claim 5. Group 3 consists of claim 6. Group 4 consists of claim 7. Group 5 consists of claim 8. Group 6 consists of claim 9. These groups of claims rise and fall independently of one another.

i. Group 1: Claims 1, 2-4, 10, 11, and 12-21

The alleged motivation cited by the Examiner for combining Sewall and Huang to reject independent claims 1 and 11, is to "accurately classify speech frames" (see Advisory Action mailed November 7, 2006). However, there are no passages indicating such motivation in either reference. For example, the Examiner quotes Sewall, col. 2, ll. 53-62 which discloses signal classification in general, not "accurate" speech frame classification. The Examiner also quotes Huang, col. 6, ll. 1-10 and 15-25, which discloses determining voiced/unvoiced segments of a random waveform <u>in the context of determining the excitation source of a</u>

synthesizer only, while remaining silent with regards to accurate representation.

Appellant asserts that the Examiner's alleged motivation is based upon Appellant's own disclosure and previous responses, and is therefore an improper use of hindsight. For Example, Appellant submits that the subject application sets forth on at least pages 2-4 that periodicity characteristics and spectral characteristics may be used to obtain "accurate" discrimination of speech and voice band data. The Examiner merely viewed the present application and intended to select prior art containing "an autocorrelation periodicity function" without citing specific evidence of motivation to combine the references, other than providing conclusory statements regarding motivation and obviousness derived from Appellant's own disclosure.

In support, the Examiner simply cites a single paragraph of Huang and Sewall (see Office Action mailed June 22, 2006, page 3) which, as stated above, in fact have nothing to do with the accurate representation of speech samples. Accordingly, absent evidence of such motivation, a *prima facie* case of obviousness under 35 U.S.C. §103(a) has not been established and the rejection must be withdrawn.

As set forth in two recent cases decided by the Court of Appeals for the Federal Circuit (CAFC), In re Dembiczak, 175 F.3d 994, 999, 50 USPQ2d 1614, 1617 (Fed.Cir. 1999) and In re Kotzab, 217 F.3d 1365, 1371, 55 USPQ2d 1313, 1317 (Fed.Cir. 2000) there are rigorous requirements for establishing a prima facie case of obviousness under 35 U.S.C. §103(a).

To establish obviousness based on a combination of elements disclosed in the prior art, there must be some motivation, suggestion, or teaching of the desirability of making the specific combination that was

made by the Appellant. The motivation suggestion or teaching may come explicitly from one of the following:

(a) the statements in the prior art (patents themselves)

(b) the knowledge of one of ordinary skill art, or in some cases,

(c) the nature of the problem to be solved. [1]

In order to establish a *prima facie* case of obviousness under 35 U.S.C. §103(a), the Examiner must provide particular findings as to why the two pieces of prior art are combinable. [2]

Appellant submits that neither Sewall nor Huang teach or suggest combining their features to arrive at independent claim 1; nor does the Examiner cite any particular passage which actually provides evidence that such a combination would be obvious to one of ordinary skill in the art. On the contrary, the disclosed references seek to overcome differing problems and therefore do not constitute an obvious combination. Sewall is directed to a **method of signal classification** to determine which type of voice-band traffic is being carried (Sewall, col. 1, lines 9-20, and col. 2, lines 22-42); Namely, classifying a signal based on a type of voice band traffic **<u>received from a communication channel</u>**. Huang is directed to an **apparatus to digitally encode voice messages** for high compression for mixing with data sent over a communications channel (Huang, col. 3, lines 10-25); Namely, separating voiced and un-voiced portions of speech in general to enable high-compression **<u>to be sent over a communication channel</u>** (Huang col. 3, ll. 15-21).

Therefore, given such distinct and differing problems solved by the references, neither reference provides any evidence, teaching or suggestion of combination with the other.

[1] See *Dembiczak,* 50 USPQ at 1614.
[2] See *Dembiczak,* 50 USPQ2d at 1617.

More particularly, Sewall, as discussed above, is directed to a method of signal classification to determine which type of voice-band traffic is being carried. Furthermore, the Examiner has admitted that Sewall does not disclose "performs a periodicity parameter in an autocorrelation function" (see Office Action dated June 22, 2006, page 2). Therefore Sewall discloses classifying signals on a communication channel without a periodicity parameter. Appellant further submits that the signals must be on the channel to be classified, as disclosed by Sewall.

Huang, as discussed above, is directed to separating voiced and un-voiced portions of speech to be sent over a communications channel. Therefore Huang discloses an apparatus sending communications. Appellant submits that the apparatus disclosed by Huang operates on signals to be sent, which is distinct and different from the signals actually on the communication channel disclosed by Sewall.

Thus, it would not have been obvious to one of ordinary skill in the art to combine the teachings of Sewall and Huang.

Furthermore, Appellant submits that even assuming *arguendo* that the two references were combinable, which Appellant does not admit, the resulting combination would not render claim 1 obvious to one skilled in the art. Even if the references were combined for the sake of argument, the combination would result in an <u>apparatus for compressing signals to be sent on a communication channel, and classifying the signals actually on the communication channel **without a periodicity parameter for classification**</u>.

As such, the combination would not disclose or suggest "[a] <u>method of discriminating speech from voice-band data in a communication network</u>, comprising... calculating a self similarity ratio value, <u>representing a periodicity characteristic</u>, and an autocorrelation

coefficient value, representing a spectral characteristic, for an input signal segment...and...determining whether said input signal segment is speech or voice-band data based on said at least one of said self similarity value and said autocorrelation coefficient value" as recited in independent claim 1 (emphasis added).

Similarly, the combination would not disclose or suggest " [an] apparatus for discriminating speech from voice-band data in a communication network, comprising... calculating means for calculating a self similarity ratio value, representing a periodicity characteristic, and an autocorrelation coefficient value, representing a spectral characteristic, for an input signal segment... and ... determining means for determining whether said input signal segment is speech or voice-band data based on said at least one of said self similarity value and said autocorrelation coefficient value" as recited in independent claim 11 (emphasis added).

Furthermore, Appellant respectfully submits that there must be some reasoning as to why one of ordinary skill would combine the references to reach the claimed invention. Sewall relies entirely on a normalized second order moment for signal classification (Sewall, col. 1, ll. 50-57, and col. 8, ll. 49-55). Sewall gives no indication or motivation to include any further parameters for signal classification nor using periodicity parameters for signal classification. Huang does not provide such reasoning either. Furthermore, as stated above, Huang is directed to signal compression, not signal classification as required by Sewall and the claimed invention. As such, the references lack such reasoning or motivation for a combination.

In view of the above arguments, Appellant asserts that the Examiner has not established the required motivation for combining the

teachings of Sewall and Huang and therefore fails to establish a *prima facie* case of obviousness under 35 U.S.C. §103(a).

Claims 2-4, 10, and 12-21 are patentable at least by virtue of their dependency upon independent claims 1 and 11.

Accordingly, Appellant respectfully requests withdrawal of the above rejections.

ii. Group 2: Claim 5

As discussed above, Appellant submits that a *prima facie* case of obviousness has not been made with regards to independent claims 1 and 11. However, even assuming *arguendo* that Sewall and Huang were combinable, which Appellant does not admit, the resulting combination would not render the limitations of claims 5-7 obvious, as further discussed below.

Neither Sewall nor Huang individually or in combination teach or suggest:

> that said input signal segment is voice-band data if said first autocorrelation coefficient is less than a <u>first autocorrelation threshold</u>, and that said input signal segment is speech if said first autocorrelation coefficient is greater than a <u>second autocorrelation threshold</u>, said second autocorrelation threshold being greater than said first autocorrelation threshold
>
> (claim 5, emphasis added)

For example, Sewall discloses signal classification. However, Sewall discloses classification based on a single threshold (see Sewall, FIG. 41A, element 100). Furthermore, Huang does not disclose autocorrelation thresholds, only autocorrelation peaks for use in compression (see Huang, FIG. 14). Furthermore, even if one could construe the use of multiple thresholds in either reference, the references <u>do not disclose these particular limitations, nor can these limitations be inherent</u>. Thus,

a *prima facie* case of obviousness cannot be made with regards to claim 5.

Appellants further submit that the Examiner has simply encapsulated the limitations of claim 5 and folded said limitations under a common rejection, without providing particular reasoning behind said rejection. Appellant submits that neither Sewall nor Huang, individually or in combination, teach or fairly suggest these limitations. Therefore, this rejection is improper and should be withdrawn.

Accordingly, Appellant respectfully requests that the above rejection be withdrawn.

iii. Group 3: Claim 6

Claim 6 sets forth that the method "calculates <u>second and third autocorrelation coefficients</u> as second and third spectral characteristic values respectively". As discussed above with regards to claim 5, neither reference discloses the use of multiple thresholds, therefore the references <u>do not disclose these limitations, nor can these limitations be inherent</u>. Thus, a *prima facie* case of obviousness cannot be made with regards to claim 6.

Appellants further submit that the Examiner has simply encapsulated the limitations of claim 6 and folded said limitations under a common rejection, without providing particular reasoning behind said rejection. Appellant submits that neither Sewall nor Huang, individually or in combination, teach or fairly suggest these limitations. Therefore, this rejection is improper and should be withdrawn.

Accordingly, Appellant respectfully requests that the above rejection be withdrawn.

iv. Group 4: Claim 7

Claim 7 sets forth that the method "determines that said input signal segment is voice-band data if a sum of said <u>second autocorrelation coefficient and said third autocorrelation coefficient is less than a fifth autocorrelation threshold</u>".

As discussed above with regards to claims 5-6, neither reference discloses the use of multiple thresholds. Furthermore, neither reference even hints at computing a sum for calculation purposes in that no more than one threshold is disclosed. Therefore, the references <u>do not disclose these particular limitations, nor can these limitations be inherent</u>. Thus, a *prima facie* case of obviousness cannot be made with regards to claim 7.

Appellants further submit that the Examiner has simply encapsulated the limitations of claim 7 and folded said limitations under a common rejection, without providing particular reasoning behind said rejection. Appellant submits that neither Sewall nor Huang, individually or in combination, teach or fairly suggest these limitations. Therefore, this rejection is improper and should be withdrawn.

Accordingly, Appellant respectfully requests that the above rejection be withdrawn.

v. Group 5: Claim 8

As discussed above, Appellant submits that a *prima facie* case of obviousness has not been made with regards to independent claims 1 and 11. However, even assuming *arguendo* that Sewall and Huang were combinable, which Appellant does not admit, the resulting combination would not render the limitations of claim 8 obvious, as further discussed below.

Claim 8 sets forth "a sequential decision logic sequence which designates input signal segments as speech during a speech state and designates input signal segments as voice-band data during a voice-band data state".

Appellant submits that Sewall and Huang, fail to even hint at a speech state or voice-band state.

For example, Sewall discloses that classification may be distinguished based on classes (see Sewall, col. 6, lines 10-23), and that these classes are based entirely on the types of modems and speech. However, there is no distinction between a state of speech or data. As set forth on page 10, starting at line 25 of the subject application, a speech state is a state in which "the classification result remains speech until subsequent classification results indicate that the speech state is erroneous". Therefore, Sewall's disclosure of classes based on modem types and speech are not speech states or data states. Furthermore, Huang lacks disclosure of any states at all.

Therefore, Sewall and Huang, alone or in combination, fail to disclose or suggest "a sequential decision logic sequence which designates input signal segments as speech during a speech state and designates input signal segments as voice-band data during a voice-band data state" as recited in claim 8.

Accordingly, Appellant respectfully requests withdrawal the above rejection.

vi. Group 6: Claim 9

Claim 9 sets forth that "sequential decision logic sequence switches from said speech state to said voice-band data state when results of said determining step for a plurality of input signal segments indicate that said speech state is erroneous".

Appellant submits that because Sewall and Huang fail to disclose or suggest a speech state or data state, the references cannot disclose or suggest switching between these states.

Accordingly, Applicants respectfully request withdrawal of the above rejection.

the remainder of this page is left blank intentionally

X. CONCLUSION:

Appellants respectfully request the Board to reverse the Examiner's rejection of claims 1-21.

The Commissioner is authorized in this, concurrent, and future replies, to charge payment or credit any overpayment to Deposit Account No. 08-0750 for any additional fees required under 37 C.F.R. § 1.16 or under 37 C.F.R. § 1.17; particularly, extension of time fees.

Respectfully submitted,

HARNESS, DICKEY & PIERCE, PLC

By ______________________
Gary D. Yacura
Reg. No. 35,416

GDY/JRS

P.O. Box 8910
Reston, VA 20195
(703) 668-8000

APPENDIX A

CLAIMS

Claims 1-21 on Appeal:

1. A method of discriminating speech from voice-band data in a communication network, comprising:

calculating a self similarity ratio value, representing a periodicity characteristic, and an autocorrelation coefficient value, representing a spectral characteristic, for an input signal segment, wherein calculating the self similarity ratio value includes calculating a plurality of different self similarity ratio values and selecting the highest one of the plurality of different self similarity ratio values as the calculated self similarity ratio value; and

determining whether said input signal segment is speech or voice-band data based on said at least one of said self similarity value and said autocorrelation coefficient value.

2. The invention as defined in claim 1, wherein said input signal segment is a frame of *N* samples.

3. The invention as defined in claim 1, wherein

said calculating step calculates a first self similarity ratio value, corresponding to a first sample delay, as a first periodicity characteristic value; and

said determining step determines that said input signal segment is voice-band data if said first self similarity ratio value is greater than a first similarity threshold.

4. The invention as defined in claim 3, wherein

said calculating step calculates a second self similarity ratio value, corresponding to a second sample delay, as a second periodicity characteristic value, said second sample delay being greater than said first sample delay; and

said determining step determines that said input signal segment is speech if said second self similarity ratio value is greater than a second similarity threshold.

5. The invention as defined in 1, wherein

said calculating step calculates a first autocorrelation coefficient as a first spectral characteristic value; and

said determining step determines that said input signal segment is voice-band data if said first autocorrelation coefficient is less than a first autocorrelation threshold, and that said input signal segment is speech if said first autocorrelation coefficient is greater than a second autocorrelation threshold, said second autocorrelation threshold being greater than said first autocorrelation threshold.

6. The invention as defined in claim 5, wherein

said calculating step calculates second and third autocorrelation coefficients as second and third spectral characteristic values respectively, and

said determining step determines that said input signal segment is voice-band data if said second autocorrelation coefficient is less than a third autocorrelation threshold or said third autocorrelation coefficient is less than a fourth autocorrelation threshold.

7. The invention as defined in claim 6, wherein

said determining step determines that said input signal segment is voice-band data if a sum of said second autocorrelation coefficient and said third autocorrelation coefficient is less than a fifth autocorrelation threshold.

8. The invention as defined in claim 1, wherein

said calculating and determining steps are performed for a plurality of input signal segments in accordance with a sequential decision logic sequence which designates input signal segments as speech during a speech state and designates input signal segments as voice-band data during a voice-band data state.

9. The invention as defined in claim 8, wherein

said sequential decision logic sequence switches from said speech state to said voice-band data state when results of said determining step for a plurality of input signal segments indicate that said speech state is erroneous, and

said sequential decision logic sequence switches from said voice-band data state to said speech state when results of said determining step for a plurality of input signal segments indicate that said voice-band data state is erroneous.

10. The invention as defined in claim 8, wherein

results of said determining step are weighted based on energy content of the corresponding input signal segment so that determination results for low energy input signal segments are given relatively low weight when determining whether to switch from said speech state to said voice-band data state or from said voice-band data state to said speech state.

11. An apparatus for discriminating speech from voice-band data in a communication network, comprising:

calculating means for calculating a self similarity ratio value, representing a periodicity characteristic, and an autocorrelation coefficient value, representing a spectral characteristic, for an input signal segment, wherein calculating the self similarity ratio value includes calculating a plurality of different self similarity ratio values and selecting the highest one of the plurality of different self similarity ratio values as the calculated self similarity ratio value; and

determining means for determining whether said input signal segment is speech or voice-band data based on said at least one of said self similarity value and said autocorrelation coefficient value.

12. The invention as defined in claim 11, wherein said input signal segment is a frame of *N* samples.

13. The invention as defined in claim 11, wherein

said calculating means calculates a first self similarity ratio value, corresponding to a first sample delay, as a first periodicity characteristic value; and

said determining means determines that said input signal segment is voice-band data if said first self similarity ratio value is greater than a first similarity threshold.

14. The invention as defined in claim 13, wherein

said calculating means calculates a second self similarity ratio value, corresponding to a second sample delay, as a second periodicity characteristic value, said second sample delay being greater than said first sample delay; and

said determining means determines that said input signal segment is speech if said second self similarity ratio value is greater than a second similarity threshold.

15. The invention as defined in 11, wherein

said calculating means calculates a first autocorrelation coefficient as a first spectral characteristic value; and

said determining means determines that said input signal segment is voice-band data if said first autocorrelation coefficient is less than a first autocorrelation threshold, and that said input signal segment is speech if said first autocorrelation coefficient is greater than a second autocorrelation threshold, said second autocorrelation threshold being greater than said first autocorrelation threshold.

16. The invention as defined in claim 15, wherein

said calculating means calculates second and third autocorrelation coefficients as second and third spectral characteristic values respectively, and

said determining means determines that said input signal segment is voice-band data if said second autocorrelation coefficient is less than a third autocorrelation threshold or said third autocorrelation coefficient is less than a fourth autocorrelation threshold.

17. The invention as defined in claim 16, wherein

said determining means determines that said input signal segment is voice-band data if a sum of said second autocorrelation coefficient and said third autocorrelation coefficient is less than a fifth autocorrelation threshold.

18. The invention as defined in claim 11, wherein

said apparatus classifies a plurality of input signal segments as being either speech or voice-band data in accordance with a sequential decision logic sequence which designates input signal segments as speech during a speech state and designates input signal segments as voice-band data during a voice-band data state.

19. The invention as defined in claim 18, wherein

said apparatus, in accordance with said sequential decision logic sequence, switches from said speech state to said voice-band data state when results of said determining means for a plurality of input signal segments indicate that said speech state is erroneous, and

said apparatus, in accordance with said sequential decision logic sequence, switches from said voice-band data state to said speech state when results of said determining means for a plurality of input signal segments indicate that said voice-band state is erroneous.

20. The invention as defined in claim 18, wherein

said apparatus weights results of said determining means based on energy content of the corresponding input signal segment so that determination results for low energy input signal segments are given relatively low weight when said apparatus judges whether to switch from said speech state to said voice-band data state or from said voice-band data state to said speech state.

21. The method of claim 1, wherein said self similarity ratio is calculated based on more than one sample.

APPENDIX B - EVIDENCE SUBMITTED UNDER CFR 1.130, 1.131, OR 1.132

None.

APPENDIX C – RELATED APPEALS AND INTERFERENCES APPENDIX

No related appeals or interferences are known.

APPENDIX D - DECISIONS RENDERED BY THE COURT OR THE BOARD IN RELATED APPEALS AND INTERFERENCES SECTION

None.